environmental microbiology

Environmental Microbiology (2013) **15**(9), 2557–2572 doi:10.1111/1462-2920.12151

Seasonality and disturbance: annual pattern and response of the bacterial and microbial eukaryotic assemblages in a freshwater ecosystem

Adriane Clark Jones,[1]* T. S. Vivian Liao,[1] Fares Z. Najar,[3] Bruce A. Roe,[3] K. David Hambright[2] and David A. Caron[1]

[1]*Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089-0371, USA.*

[2]*Program in Ecology and Evolutionary Biology, Department of Biology, and* [3]*Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK 73019, USA.*

Summary

High-throughput pyrosequencing of SSU rDNA genes was used to obtain monthly snapshots of eukaryotic and bacterial diversity and community structure at two locations in Lake Texoma, a low salinity lake in the south central United States, over 1 year. The lake experienced two disturbance events (i) a localized bloom of *Prymnesium parvum* restricted to one of the locations that lasted from January to April, and (ii) a large (17 cm), global rain event in the beginning of May, overlaid onto seasonal environmental change. Eukaryotic species richness as well as both eukaryotic and bacterial community similarity exhibited seasonal patterns, including distinct responses to the rain event. The *P. parvum* bloom created a natural experiment in which to directly explore the effects of an Ecosystem Disruptive Algal Bloom (EDAB) on the microbial community separated from seasonal changes. Microbial species richness was unaffected by the bloom, however, the eukaryotic community structure (evenness) and the patterns of both eukaryotic and bacterial community similarity at bloom and non-bloom sites were statistically distinct during the 4 months of the bloom. These results indicate that physical and biological disturbances as well as seasonal environmental forces contribute to the structure of both the eukaryotic and bacterial communities.

Received 31 October, 2012; accepted 28 April, 2013. *For correspondence. E-mail adrianej@usc.edu; Tel. (213) 821 1800; Fax (213) 740 8123.

Introduction

Microbes (bacteria, archaea, and microbial eukaryotes) form the foundation of food webs in aquatic ecosystems, playing a variety of critical biogeochemical roles such as primary producers, nutrient recyclers, and links to higher trophic levels (Caron and Countway, 2009; Fuhrman, 2009). Natural microbial assemblages are typically composed of relatively few common taxa that perform the majority of ecological roles at any given time, yet incredible taxonomic richness is maintained in virtually every microbial community even in extreme environments (Amaral-Zettler *et al.*, 2002; Edgcomb *et al.*, 2002; Lopez-Garcia *et al.*, 2003). Rare taxa comprise most of the species richness, and have collectively become known as the microbial rare biosphere (Sogin *et al.*, 2006). There has been considerable speculation and debate recently regarding how these rare taxa are maintained in natural assemblages, and their potential to contribute to ecosystem function in an environment (Pedros-Alio, 2006; Caron and Countway, 2009).

Microbial community diversity and composition in a given location is structured and influenced by both regional (geography) and local processes such as species-to-species interactions, or environmental differences (reviewed in Lindström and Langenheder, 2011 and Martiny *et al.*, 2006). These changes can be global and repeating (e.g. in response to diurnal and seasonal cycles) (Kent *et al.*, 2007; Gilbert *et al.*, 2009) or localized (e.g. in response to environmental gradients (Bouvier and del Giorgio, 2002; Langenheder *et al.*, 2006).

Disturbance events encompass episodic extremes in these structuring forces and can be biological (introduced or eliminated species), chemical (addition of nutrients or pollutants), or physical (severe weather events, or advection) in nature. Depending on their frequency and magnitude, disturbance events can result in completely altered ecosystems (Estes *et al.*, 2011), or act as mechanisms for maintaining biological diversity and ecosystem function within an environment (Connell, 1978). The consequences and properties of disturbance have been well documented within macro-organismal communities, and are becoming more widely recognized as important factors affecting the structure of aquatic microbial

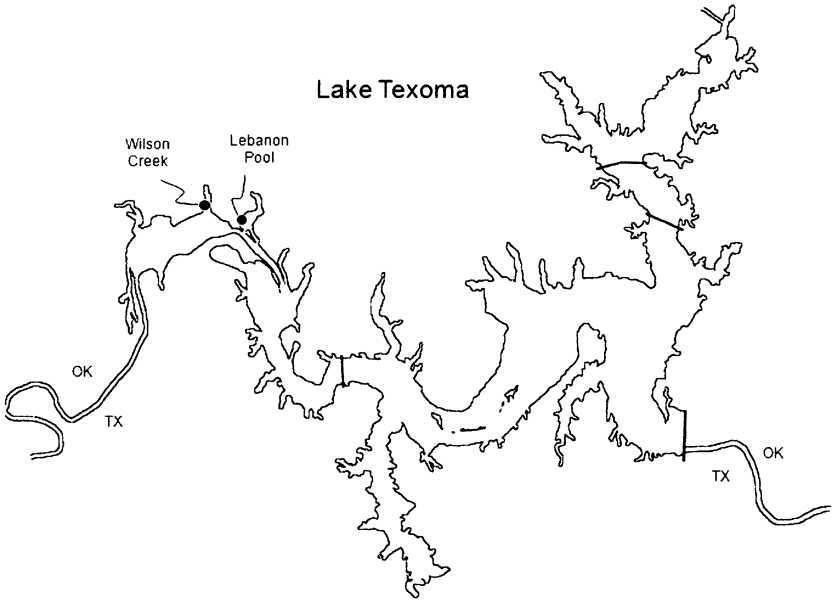

Fig. 1. Map of Lake Texoma, south central USA with the location of the two sampling sites: Wilson Creek and Lebanon Pool.

communities (Reynolds *et al*., 1993; Floder and Sommer, 1999; Allison and Martiny, 2008; Shade *et al*., 2012). Our present inability to predict changes in species composition and biogeochemical processes make observational studies of natural communities highly insightful (Fuhrman, 2009). Harmful algal blooms (HABs; the dominance of a single algal taxon possessing toxic or noxious properties) are well-documented disruptions of planktonic ecosystems that result in the disruption or collapse of aquatic food webs (Buskey *et al*., 2001; Gobler *et al*., 2005). The detrimental effects of these blooms have been well documented at higher trophic levels, including impacts on zooplankton, invertebrates, fish, birds and mammals (Scholin *et al*., 2000), yet there is limited knowledge of the effects of these events on the structure of the microbial communities (Michaloudi *et al*., 2008; Vigil *et al*., 2009).

Prymnesium parvum is a haptophyte alga that forms harmful algal blooms in brackish waters around the world, including the United States where such blooms appear to be an emerging issue. Blooms of *P. parvum* were first detected in Lake Texoma, south central USA during 2004 and this species has since formed annual blooms that can last for months (Roelke *et al*., 2010). The factors contributing to and controlling blooms of *P. parvum* are not well known, but this species does have several physiological characteristics that equip it to grow quickly and in a variety of habitats. It has a wide salinity tolerance allowing it to establish in brackish waters. *P. parvum* also produces a suite of toxins that negatively affect gill breathing animals and can lead to massive fish kills in the environment. In the laboratory, these toxins have been shown to effect zooplankton grazers and other members of the plankton (Fistarol *et al*., 2003; Skovgaard and Hansen, 2003; Evardsen and Imai, 2006). *P. parvum* is also mixotrophic, combining phototrophy with a well-developed ability to capture, subdue and ingest a wide variety of microbial species. This physiological duality presumably provides an ecological advantage relative to other algae and heterotrophic protists, and thus may play a role in bloom formation and maintenance (Tillmann, 1998; 2003). Large-scale HAB events such as the ones formed by *P. parvum* constitute major biological disturbance events in pelagic ecosystems that can dramatically affect food web structure and function and have been termed Ecosystem Disruptive Algal Blooms (EDABs) (Sunda *et al*., 2006; Gobler and Sunda, 2012).

Eukaryotic and bacterial community structure was examined in monthly samples collected at two locations in Lake Texoma, Oklahoma, USA for 1 year. Lake Texoma (Fig. 1) is located along the Texas and Oklahoma border. Salinity in the lake is generally high. Practical salinity units (PSU) averaged ~ 0.8 between from 1959 to 2008, and values can range from 0.5 to 2.5 PSU (Hambright *et al*., 2010). Cyclical patterns in eukaryotic species richness as well as eukaryotic and bacterial community similarity were observed at both locations that indicated seasonal environmental forcing. Two episodic disturbances were overlaid on this seasonal cycle; a large spring rain event at both locations in May and a bloom of *P. parvum* at one of the sampling locations but not the other during the winter. These events resulted in major changes in microbial community structure and community similarity that were comparable in magnitude or greater than responses brought

about by seasonal changes. Despite these major shifts in microbial community structure to short-term, environmental or biological disturbances, community composition returned to a cyclical annual pattern within the 1-year study.

Results

Annual patterns of temperature, salinity, pH, dissolved oxygen, nutrients, chlorophyll, and cell abundances

Trends in the physical, and chemical parameters of Lake Texoma reflected the seasonality and two disturbance events of the system: (i) Lebanon Pool experienced a 4 month bloom of the alga *Prymnesium parvum* (Fig. 2H), and (ii) both locations experienced a large (17 cm) and anomalous rain event on 3 May 2009 (Fig. S1) (McPherson *et al*., 2007).

Lebanon Pool and Wilson Creek exhibited similar patterns of surface water temperature, salinity, pH, dissolved oxygen, and dissolved nutrients over the length of the study period (Fig. 2A–F). Water temperature at both locations (Fig. 2A), had winter minima of approximately 5°C during January followed by progressive warming to summer maxima of approximately 30°C during July, and subsequent cooling through the fall. Seasonal changes in temperature at both locations displayed good fit ($r^2 = 0.8$) to sinusoidal regressions with set periods of 1 year. The patterns of salinity, pH and dissolved oxygen measurements each had the highest values from December to April, precipitously decreased with the rain event in May (Fig. S1), and then gradually increased through the summer and fall (Fig. 2B–D). Salinities (Fig. 2B) at both sites, ranged from 2.5 to 3 PSU and were relatively constant from November through April. Salinities decreased to < 0.5 PSU in May, gradually increased through September, and were approximately 1 PSU from August to October. Dissolved nitrogen (Fig. 2E solid lines) at the two locations had similar ranges (25–75 μM) and fluctuations throughout the year. Dissolved phosphorous concentrations (Fig. 2F solid lines) in the two sites were mostly constant (1–2 μM) throughout the year, with two major increases (to 3–5 μM) in May and September. Total nitrogen (Fig. 2E dotted lines) ranged from 75 to 100 μM in Lebanon Pool, and from 75 to over 150 μM in Wilson Creek. Total phosphorous (Fig. 2F dotted lines) ranged from 4 to 10 μM in Lebanon Pool, and from 4 to 15 μM in Wilson Creek. Trends in chlorophyll *a* (Fig. 2G) were similar at both study sites, but were not identical. Month-to-month changes in chlorophyll concentration in Lebanon Pool were highly variable but overall showed a modest increase (from ~ 50 to ~ 90 μg l^{-1}) from late-January to April. Chlorophyll values in Wilson Creek exhibited progressive increases (maximum of 120 μg l^{-1}) from mid-March to mid-April. Chlorophyll concentrations at both locations decreased during May to approximately 4 μg l^{-1}, coinciding with the timing of the rain event (Fig. S1) and the decreased values of salinity, pH and dissolved oxygen at the two locations (Fig. 2B–D). Chlorophyll concentrations at both sites increased steadily from June onward and reached maxima of 160 and 190 μg l^{-1} during early fall (Fig. 2G).

Microscopical counts of *P. parvum* (Fig. 2H) differed between the two study sites. Lebanon Pool experienced a large bloom of *P. parvum* during winter-spring. Cell abundances increased steadily from January to mid-February at this site, sustained maximal abundances of $1.5–2 \times 10^5$ cells ml^{-1} into April, and then decreased rapidly from April to May. The demise of the *P. parvum* bloom began prior to the rain event at the beginning of May (Fig. S1). The *P. parvum* bloom coincided with a relatively modest (and highly variable) increase in chlorophyll concentrations during the winter in Lebanon Pool (Fig. 2G). *P. parvum* cells were detected at Wilson Creek during the same period of time, but never exceeded 800 cells ml^{-1}. The haptophyte was undetected by routine microscopical examination at both locations from May until the end of the study (Fig. 2H inset).

Genetic signatures of Prymnesium *cells*

The bloom of *P. parvum* in Lebanon Pool that was documented with microscopical counts (Fig. 2H) was also apparent in the 18S and 16S ribosomal sequence data. The 18S rDNA sequences from Lebanon Pool identified as *Prymnesium* (Fig. 3A) comprised 30–50% of the total number of sequences for the months of January, February, March and April. Tags identified as *Prymnesium* in Wilson Creek during the same time period were detected but constituted no more than 2% of the total number of sequences in those samples, consistent with the low number of cells detected by microscopy (Fig. 2H inset). Only low relative abundances of *Prymnesium* were detected at both sampling locations from June to October (Fig. 3A inset). 16S rDNA sequences identified as *Prymnesium* plastids (Fig. 3B) comprised approximately 20% of the total number of sequences for the months of January, February, March and April, in Lebanon Pool, whereas sequences were at low abundances (approximately 2% of total sequences) in Wilson Creek.

Bacterial and eukaryotic operational taxonomic units

Operational taxonomic unit (OUT) calling at 97% similarity was performed on the subsampled 18S and 16S (plastids removed) datasets (see *Experimental procedures*). The eukaryotic 18S dataset contained 9770 total OTUs and the bacterial 16S dataset contained total 8392 OTUs (Table 1). Singletons (OTUs represented by a single

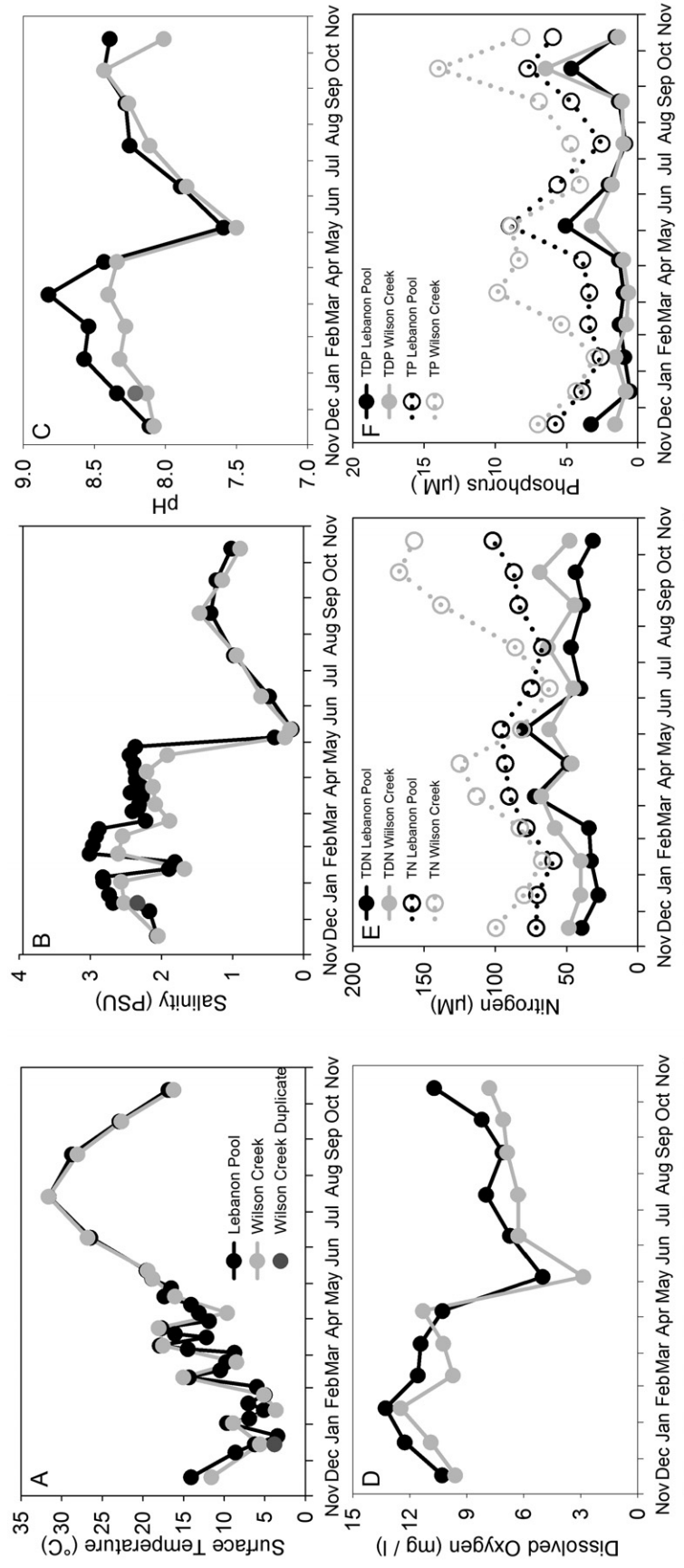

Fig. 2. An annual cycle at Lebanon Pool (black circles) and Wilson Creek (grey circles) in Lake Texoma of surface water temperature (A), salinity (B), pH (C),dissolved oxygen (D), dissolved nitrogen (E – solid symbols) and total nitrogen (E – open symbols), dissolved phosphorous (F – solid symbols) and total phosphorous (F – open symbols), extracted chlorophyll a, (G) and *Prymnesium parvum* cell abundances determined by microscopical counts (H). Data and samples were collected from November 2008 to October 2009. A duplicate sample collected in Wilson Creek during December 2008 is shown as a dark grey circle in (A, B, C, G and H). The inset in (H) illustrates the same data with a narrower range of the *y* axis showing low but detectable abundances of *P. parvum* in Wilson Creek during late winter and spring.

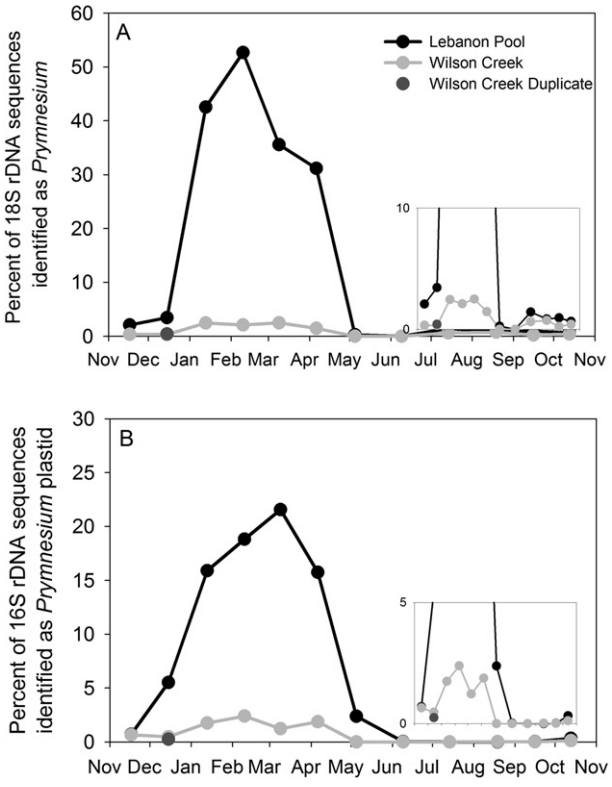

Fig. 3. The percent contribution of DNA sequences in samples from Lebanon Pool (black circles), and Wilson Creek (grey circles) in Lake Texoma attributed to genus *Prymnesium.* The percent contribution of 18S sequences attributed to *Prymnesium* (A), and percent contribution of 16S sequences attributed to a *Prymnesium* plastid (B). Samples were collected from November 2008 to October 2009. A duplicate sample collected in Wilson Creek during December 2008 is shown as a dark grey circle. The insets illustrate the same data with a narrower range of the *y* axis showing low but detectable abundances of *Prymnesium* sequences in Wilson Creek during late winter and spring.

sequence) comprised 59% (5753) of the 18S OTUs, and 3% of the total number of sequences. Singletons in the 16S dataset comprised 64% (5401) of the total number of OTUs and 6% of the total number of sequences (Table 1).

The numbers of observed OTUs per month during the winter months of January to April were lower than the numbers observed during the rest of the year in both datasets (18S and 16S) at both sampling sites (Table 1; Fig. 4A and B). The number of 18S OTUs observed each month declined from November to January, and remained low (250–327 per sample for Lebanon Pool and 425–539 per sample for Wilson Creek) through April. The numbers of eukaryotic OTUs detected in Lebanon Pool during the *P. parvum* bloom (Fig. 2H), (January to March) were 1/3 the number of eukaryotic OTUs detected in Wilson Creek during the same period. Both sites experienced marked increases in the number of observed OTUs during May, reached maxima during August (~ 1200) and declined steadily through the fall. The number of eukaryotic OTUs observed each month in both Lebanon Pool and Wilson Creek showed a yearly seasonal pattern, with good agreement to a sine wave function ($r^2 = 0.8$) with a set period of 1 year (Fig. 4A).

The numbers of 16S OTUs per month in both locations were lowest from January to April, but otherwise the values remained relatively high (approximately 750) and constant (Table 1; Fig. 4B). The lowest number of OTUs observed in each site occurred in Lebanon Pool during February (400) and Wilson Creek during January (394). The largest month-to-month increases in the numbers of bacterial OTUs occurred from April to May at both study sites, similar to trends for the 18S data. However, the largest numbers of OTUs per sample were observed in Lebanon Pool during May (871) and Wilson Creek during June (911) (Table 1; Fig. 4B).

The dominant 18S OTU in Lebanon Pool for the four months during the *P. parvum* bloom was identified as a Prymnesiales. Sequences in the Prymnesiales OTU constituted 29–49% of the total sequences for those four months (Table 1). This degree of dominance of a single OTU was not observed at any other time in the study. The dominant bacterial OTU during the *P. parvum* bloom in Lebanon Pool was a freshwater *Synechococcus* in 3 of the 4 months (Table 1).

Diversity of 18S and 16S OTUs

The representative trends in the Chao 1 (Fig. 4C) and Catchall (Fig. 4E) species richness estimates for the 18S rDNA data were similar for the two sampling locations (mostly overlapping confidence intervals). The richness estimators were lowest (< 750 and < 2500 for Chao I and Catchall respectively) during the winter-spring months of January to April and highest (~ 3000 and ~ 7000 respectively) during the summer months of June and July (Fig. 4C and E). Both richness estimators showed a yearly seasonal pattern with good agreement to a sine wave function ($r^2 = 0.8$) with a set period of 1 year. These changes also reflected the trends in the observed 18S OTUs (Fig. 4A), and the two richness estimators had excellent correlations (spearman correlations > 0.93 (*P*-values < 0.001) with the observed 18S richness at both locations. The strong dominance of the Prymnesiales OTU in Lebanon Pool during the months of January to April (Table 1) did not cause a significant difference between the richness estimates obtained at the bloom (Lebanon Pool) and non-bloom (Wilson Creek) sites (Fig. 4C and E).

The Chao 1 and Catchall values for bacterial richness (Fig. 4D and F) indicated a trend towards higher values during the summer at both sampling locations, but these differences were not significant because of the high

Table 1. A summary of eukaryotic 18S and bacterial 16S rDNA sequences.

	18S Lebanon Pool				18S Wilson Creek			
			Top OTU				Top OTU	
Month	# OTUs	# Singletons[a]	% Sequences	Taxonomic ID[b]	# OTUs	# Singletons[a]	% Sequences	Taxonomic ID[b]
November	646	383	28%	OTU 10 Stramenopile; Chrysophyte	753	430	14%	OTU 22 Alveolata; Ciliate
December	526	280	8%	OTU 47 Alveolate; Ciliate	630	363	20%	OTU 3 Katablepharid; Katablepharidaceae
Duplicate					661	386	19%	OTU 3 Katablepharid; Katablepharidaceae
January	296	155	43%	OTU 1 Haptophyceae; Prymnesiales	454	216	21%	OTU 5 Alveolate; Ciliate
February	326	201	53%	OTU 1 Haptophyceae; Prymnesiales	539	300	16%	OTU 14 Stramenopile; Chrysophyte
March	253	162	35%	OTU 1 Haptophyceae; Prymnesiales	482	241	12%	OTU 8 Metazoa; Rotifer
April	328	210	31%	OTU 1 Haptophyceae; Prymnesiales	425	228	13%	OTU 15 Haptophyte; Pavlovales
May	966	596	12%	OTU 9 Rhizaria; Cercozoa	788	401	11%	OTU 9 Rhizaria; Cercozoa
June	1004	610	11%	OTU 20 Alveolate; Ciliate	998	614	22%	OTU 6 Viridiplante; Chlorophyte
July	1212	666	7%	OTU2 Stramenopile; Diatom	1208	674	6%	OTU 2 Stramenopile; Diatom
August	1021	548	9%	OTU2 Stramenopile; Diatom	1067	589	8%	OTU 2 Stramenopile; Diatom
September	877	503	8%	OTU 41 Stramenopile; Diatom	912	510	11%	OTU 11 Stramenopile; Diatom
October	812	466	13%	OTU 29 Stramenopile; Chrysophyte	910	530	8%	OTU 24 Alveolate; Ciliate
Lake[c]	9770	5753						

	16S Lebanon Pool				16S Wilson Creek			
			Top OTU				Top OTU	
Month	# OTUs	# Singletons[a]	% Sequences	Taxonomic ID[b]	# OTUs	# Singletons[a]	% Sequences	Taxonomic ID[b]
November	741	443	3%	OTU 10 Verrucomicrobia; Spartobacteria	729	441	5%	OTU 1 Alphaproteobacteria; SAR11 clade-LD12
December	726	463	9%	OTU 37 Proteobacteria; Gammaproteobacteria	602	390	9%	OTU 1 Alphaproteobacteria; SAR11 clade-LD12
Duplicate					491	251	10%	OTU 1 Alphaproteobacteria; SAR11 clade-LD12
January	435	252	8%	OTU 3 Cyanobacteria; *Synechoccous*	394	229	9%	OTU 6 Bacteroidetes; Flavobacteria
February	400	224	24%	OTU 7 Bacteroidetes; Flavobacteria	518	339	7%	OTU 2 Verrucomicrobia; Spartobacteria
March	512	293	8%	OTU 3 Cyanobacteria; *Synechoccous*	639	428	9%	OTU 2 Verrucomicrobia; Spartobacteria
April	552	341	10%	OTU 3 Cyanobacteria; *Synechoccous*	520	338	9%	OTU 8 Alphaproteobacteria; SAR11 clade-LD12
May	871	516	5%	OTU 65 Bacteroidetes; Flavobacteria	797	424	4%	OTU 30 Bacteroidetes; Sphingobacteria
June	714	450	7%	OTU 4 Actinobacteria; Actinobacteria	911	594	8%	OTU 1 Alphaproteobacteria; SAR11 clade-LD12
July	735	460	7%	OTU 1 Alphaproteobacteria; SAR11 clade-LD12	746	464	13%	OTU 1 Alphaproteobacteria; SAR11 clade-LD12
August	768	460	6%	OTU 4 Actinobacteria; Actinobacteria	737	450	7%	OTU 2 Verrucomicrobia; Spartobacteria
September	854	510	8%	OTU 4 Actinobacteria; Actinobacteria	896	580	5%	OTU 2 Verrucomicrobia; Spartobacteria
October	781	449	5%	OTU 4 Actinobacteria; Actinobacteria	810	510	3%	OTU 2 Verrucomicrobia; Spartobacteria
Lake[c]	8392	5401						

a. Singletons are defined as OTUs with only one sequence per that sample.
b. Taxonomic ID based on 95% or greater sequence similairty to SILVA or NCBI reference database.
c. Lake values are integrated at both sites over the 12 months (25 discrete samples).
OTUs and singletons established at 97% similarity, organized by month (November 2008 to October 2009) and location (Lebanon Pool and Wilson Creek). Sequences attributed to the plastids of eukaryotic taxa have been removed. All samples were normalized to the same number of sequences (18S = 7678 seqs/month or 191 950 total) (16S = 3597 seqs/month or 89 925 total) prior to calling OTUs.

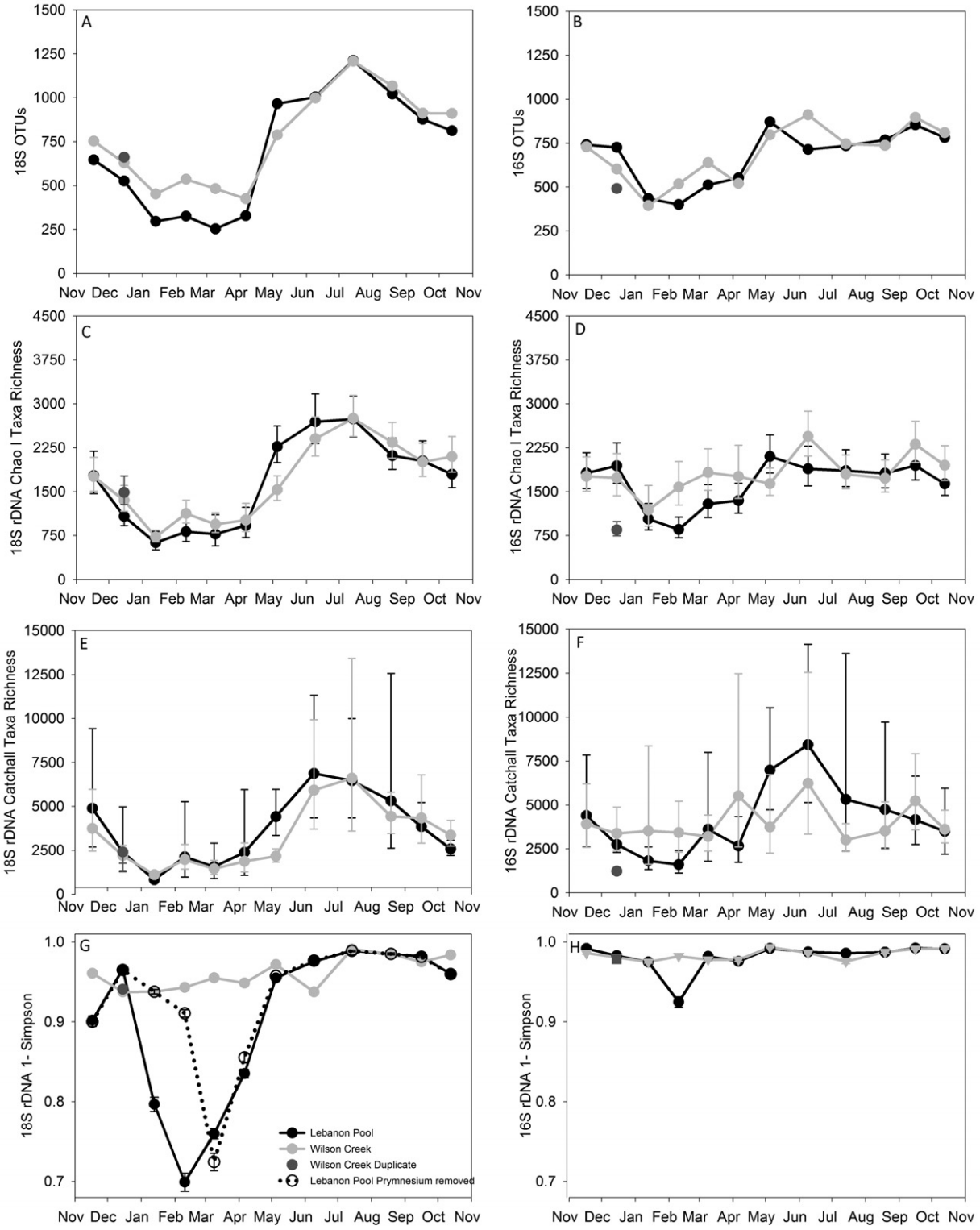

Fig. 4. Diversity indices determined over an annual cycle (November 2008 to October 2009) in Lebanon Pool (black circles), Wilson Creek (grey circles) and a duplicate sample from Wilson Creek in December (dark grey circles) for eukaryotic OTUs (A, C, E, G) and bacterial OTUs (B, D, F, H). Indices included the observed number of OTUs in each sample (A and B), the Chao 1 estimation of species richness (C and D), the CatchAll estimation of species richness (E and F), and the 1-Simpson Index of Diversity estimator (G and H). Plastid sequences were removed from the bacterial dataset prior to calculating diversity estimates. Error bars represent the computed upper and lower 95% confidence intervals. The dotted line in (G) represents the 1-Simpson Index of Diversity estimate for the Lebanon Pool dataset with *Prymnesium* OTUs removed and the data re-analysed.

variances. Species richness in Lebanon Pool displayed the lowest values (~ 750 and < 2000 for Chao 1 and Catchall respectively) during January and February and highest values during May or June (~ 2250 and ~ 7000 respectively; Fig. 4D and F black lines). Values of estimated species richness for Wilson Creek fluctuated from approximately 800 to 2250 (Chao 1 index; Fig. 4D) and from 3000 to 6000 (Catchall index; Fig. 4F) throughout the year. The Chao 1 estimator (Fig. 4D) showed good correlations (spearman values of 0.75 and *P*-values < 0.01) to the observed OTU richness at both locations (Fig. 4B). The catchall richness estimators did not correlate well to the observed 16S richness.

The annual patterns for the Simpson Index of Diversity for the eukaryotic community (Fig. 4G) was markedly different between the two locations for part of the year. This index was significantly lower (~ 0.7–0.8) for Lebanon Pool samples collected during the months of January to April compared with values obtained for the rest of the year in Lebanon Pool and also compared with all of the values obtained for Wilson Creek (values > 0.9). The timing of the low values coincided with the *P. parvum* bloom in Lebanon Pool (Fig. 2H). The index was relatively constant throughout the year for Wilson Creek samples (average = 0.96) (Fig. 4G). When sequences attributed to the Prymnesiales OTU were removed from the dataset, the sequences subsampled again for normalization, and the Simpson index recalculated (Fig. 4G dotted line), the 1-Simpson values were still statistically lower in Lebanon Pool compared with the non-bloom site of Wilson Creek in 3 (February, March and April) of the 4 months in which the haptophyte bloomed (Figs 2H, 3A and B).

Values of the Simpson Index determined for the bacterial assemblages (Fig. 4H) were remarkably similar, constant and high (0.98) at both locations throughout the year-long study. The one exception to this generality was the February sample in Lebanon Pool, which was characterized by a single, abundant bacterial OTU (Flavobacteria) that constituted 24% of all sequences (Table 1).

The true replicates (two separate samples from Wilson Creek during December obtained on a single sampling date, and processed in parallel) yielded similar values for the total number of OTUs, singletons and taxonomic assignment of OTUs for the 18S dataset and 16S datasets (Table 1), and similar 18S Chao 1, 18S Catchall and 18S and 16S Simpson Diversity estimators (Fig. 4 C, E, G and H).

Seasonal and site-specific patterns in microbial community similarities

Community similarities built from Bray–Curtis and Jaccard indices for the eukaryotic and bacterial assemblages were examined across all months, and between both sampling sites. The Bray–Curtis and Jaccard similarity values for the eukaryotic and bacterial assemblages revealed a cyclical pattern of change over the course of the year at both sites, as well as site-specific differences during a portion of the year. The similarity matrices built from the Bray–Curtis relative abundance measures and the Jaccard presence/absence measures were similar (Spearman correlations of > 0.9) for the 18S and 16S datasets. The average Bray–Curtis similarity across all monthly samples at each location was low (< 8% for the 18S data and < 18% for the 16S data) (Fig. S2A and C). The global analysis of variance test (ANOSIM) confirmed differences between the months: (*R*-test statistics of 0.53 and *P*-value = 0.002 for the 18S data, and 0.79 and *P*-value = 0.001 for the 16S data). The average Jaccard similarity across the year at each location also was low (< 9% for the 18S data, and < 9% for the 16S data) (Fig. S2B and D). The ANOSIM test confirmed the differences between the months: (R test statistics of 0.73 and *P*-value = 0.001 for the 18S data, and R test statistic of 0.75 and *P*-value = 0.001 for the 16S data). The eukaryotic or bacterial assemblages between the two locations (Lebanon Pool and Wilson Creek) could not be distinguished with a global ANOSIM test when all months were considered together. However, when the samples were grouped based on months with high or low cell abundances of *P. parvum* (see *Experimental procedures*), there was a strong and significant difference observed between the two locations [R test statistics of 0.98 (Bray–Curtis) and 1 (Jaccard) for the 18S data, and R test statistics of 0.80 (Bray–Curtis) and 0.88 (Jaccard) for the 16S data; (*P*-values < 0.03)]. In both the 18S and 16S datasets we did not detect a difference between the locations for the months of November, December and May to October.

The two-dimensional MDS (multidimensional scaling) representations of the eukaryotic and bacterial Bray–Curtis (Fig. 5A and C) and the Jaccard similarity matrices (Fig. 5B and D) each had low stress values (≤ 0.10) indicating good spatial positioning of the samples. The strong site-specific differences detected with the ANOSIM test for the months of January to April (the period of the *P. parvum* bloom in Lebanon Pool) were conspicuous on the MDS plots. These separations were apparent when relative abundance (Bray–Curtis; Fig. 5A and C) or presence/absence (Jaccard; Fig. 5B and D) were used to calculate the similarities. The 18S and 16S MDS arrangements of the samples from Lebanon Pool and Wilson Creek for the month of May (the time of the rain event) were similar and also spatially distinct from all other samples. The eukaryotic or bacterial communities at the two locations remained closely grouped from July through October and returned to the proximity of the November

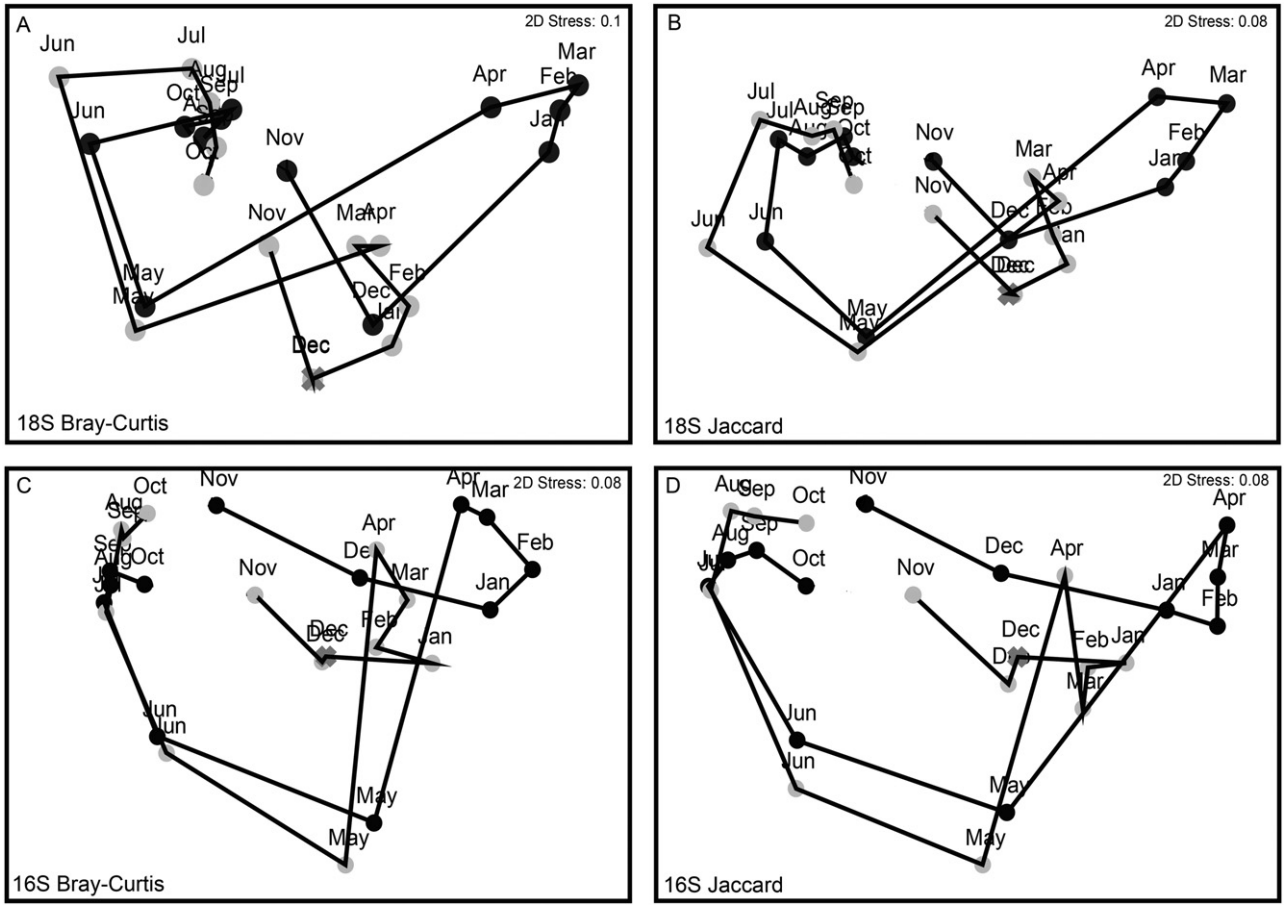

Fig. 5. Annual patterns of eukaryotic (A and B) and bacterial (C and D) community structure in Lebanon Pool (black circles), Wilson Creek (grey circles) and a duplicate sample from Wilson Creek in December (grey x) based on two-dimensional non-metric multidimensional scaling representation of Bray–Curtis similarity values (A and C) and Jaccard similarity values (B and D) calculated from non-transformed absolute OTU abundances. Sequential months have been linked with lines to indicate annual trajectories of the samples. Spearman correlations (*P*-values = 0.001) to a 1-year cyclical model matrix are as follows: 18S Bray–Curtis, L.P = 0.53 and W.C = 0.54; 18S Jaccard, L.P. = 0.62 and W.C. = 0.59; 16S Bray–Curtis, L.P. = 0.64 and W.C. = 0.52; and 16S Jaccard, L.P. = 0.70 and W.C. = 0.62).

samples in the previous year (i.e. the beginning of the study period). A statistical test (see *Experimental procedures*) for annual cyclicity of the 18S and 16S datasets at both Lebanon Pool and Wilson Creek showed good Spearman correlations to a model cyclical matrix (0.52 to 0.70 with *P*-values = 0.001) indicating that these communities experienced a similar seasonal progression throughout the year (Fig. 5A–D).

The placement of the eukaryotic and bacterial communities in the MDS plots were spatially similar for the Bray–Curtis (Fig. 5A vs. C) and Jaccard indices (Fig. 5B vs. D). This similarity was statistically supported with strong Spearman correlations between the 18S and 16S Bray–Curtis similarity matrices (0.73 *P*-value = 0.001) and between the 18S and 16S Jaccard similarity matrices (0.87 *P*-value = 0.001).

The duplicate samples collected from Wilson Creek during December and processed in parallel were highly similar and not statistically different on the basis of the SIMPROF test. Bray–Curtis similarity values for the two duplicate samples were 85% for the 18S and 75% for the 16S datasets. Jaccard values for the two duplicate samples were 30% for the 18S and 23% for the 16S datasets (Fig. S2). The two duplicate samples occupied close physical space on the MDS plots (Fig. 5A–D). The duplicate samples also provided an experimental maximum of the Bray–Curtis and Jaccard similarity measures. No two 18S samples were more similar than the values for the duplicate December samples of 85% or 30% (Fig. S2A and B), no two 16S Bray–Curtis values were more similar than 75% (Fig. S2C), and only 4 pairs of Jaccard values were equal to or greater than 23% similar (Fig. S2D). Complete hierarchical cluster diagrams based on group average scores for the Bray–Curtis and Jaccard similarities for the 18S and 16S datasets are presented in Fig. S2A–D).

The two BEST analyses returned the following sets of environmental variables that best correlated with the

detected patterns in 18S or 16S community similarity for the entire dataset (both locations). The 18S communities were best correlated (P of 0.844, *P*-value = 0.01) with temperature, salinity, and *P. parvum* cell abundances. The 16S community patterns were best correlated (P of 0.738, *P*-value = 0.01) with temperature, salinity, dissolved oxygen, pH, and *P. parvum* cells. When the biological variables were removed the 18S communities were best correlated (P of 0.663, *P*-value = 0.01) with temperature, salinity, and pH, and the 16S community patterns were best correlated (P of 0.661, *P*-value 0.01) with temperature, salinity, and pH.

Discussion

High-throughput sequencing of ribosomal DNA fragments was applied in this study to characterize changes in the eukaryotic and bacterial assemblages of a freshwater ecosystem over an annual cycle. Pyrosequencing has been increasingly employed to examine species richness and the relative abundances of microbial taxa in natural samples from a wide variety of ecosystems (Amaral-Zettler *et al*., 2009; Gilbert *et al*., 2009; Behnke *et al*., 2011; Edgcomb *et al*., 2011).

We normalized our data to avoid artefacts arising from estimating diversity indices from libraries containing different numbers of sequences (Gihring *et al*., 2012),and obtained good replication of our sequencing approach, as well as good agreement between traditional and molecular approaches. The true duplicate samples collected in December displayed similar numbers of OTUs, estimates of diversity and community relatedness for 18S and 16S datasets, and we obtained good agreement between absolute abundances of *P. parvum* by microscopical counts and relative abundances of sequences attributed to the alga. The timing and duration of the high abundances of 18S and 16S rDNA sequences attributed to *Prymnesium* (Fig. 3A and B) were consistent with the patterns of cell abundances of *P. parvum* obtained by microscopical examination for the bloom months (Fig. 2H). Further, each location had excellent Spearman correlations [0.90 and 0.89 in Lebanon Pool (*P*-values << 0.001) and 0.85 and 0.83 in Wilson Creek (*P*-values << 0.001)] between relative abundances of *P. parvum* determined by rDNA sequences and cell counts. Although our approach cannot address some inherent caveats regarding the use of sequence information for ecological investigations (e.g. variations in gene copy number, PCR bias, etc.) (Kunin *et al*., 2010; Medinger *et al*., 2010), pyrosequencing greatly improved our ability to detect taxa present at low abundances in samples. This approach allowed us to characterize a greater percentage of the microbial community of Lake Texoma, for example: *Prymnesium* affiliated 18S rDNA sequences were detected at low percentages although the alga was not detected by microscopical observation (Fig. 3A compared with 2H).

Seasonal and episodic features of Lake Texoma

The absolute values, as well as the seasonal and episodic trends of temperature, salinity, nutrients, *P. parvum* cell counts and chlorophyll concentrations observed throughout this year-long study period (November 2008 to October 2009) were representative of these parameters at littoral sites within the lake (Hambright *et al*., 2010). Seasonal trends in surface water temperature (Fig. 2A) reflect the strong overall environmental seasonality of the system. Total nitrogen and phosphorous values were high and qualify the lake for eutrophic status (Wetzel, 2001). Spring rain events are stochastic in nature and have, in the past, resulted in large-scale and rapid replacement of lake water in its littoral environments (Hambright *et al*., 2010). Both sampling locations experienced a dramatic rain event in this study (Fig. S1) which resulted in related environmental perturbations including decreases in salinity, pH and dissolved oxygen (Fig. 2B–D) in the beginning of May. Localized blooms of *P. parvum* in Lebanon Pool have become annually repeating biological features of Lake Texoma since 2004, yet, Wilson Creek has not typically experienced blooms (Hambright *et al*., 2010; Zamor *et al*., 2012). We documented one of the largest blooms to date of *P. parvum* ($> 10^5$ cells ml^{-1}) in Lebanon Pool that lasted from January to April. *P. parvum* cells were present in Wilson Creek during this study but abundances were generally 100-fold lower than in Lebanon Pool (Fig. 2H and Fig. 3A and B). Together, these features (temperature, salinity, pH, dissolved oxygen, nutrients, and *P. parvum* cell abundances) set the environmental context for exploring the combined importance of gradual seasonal forcing and episodic events (massive rain events and a harmful algal bloom) in shaping the microbial communities of the lake. Statistical analyses suggested that the eukaryotic and bacterial similarity patterns were driven by a combination of the seasonally changing environmental variable of temperature, and disturbance-related variables including salinity and *P. parvum* cell counts.

Response of the microbial community to seasonal environmental forcing

The observed and predicted eukaryotic species richness at both locations exhibited cyclical trends over the year with highest diversity in the summer and lowest in the winter (Fig. 4A, C and E). The data were collected for just 1 year but the annual pattern was similar for both locations which suggests that eukaryotic taxa richness

responds directly to seasonally fluctuating environmental parameters such as day length or water temperature (Sommer *et al*., 1986).The annual patterns in the observed and predicted numbers of bacterial OTUs were similar at the two locations, but fluctuations did not exhibit a strong cyclical pattern (Fig. 4B, D and F). This suggests that bacterial taxonomic richness, as measured in this study, was less sensitive to seasonal forcing factors than the eukaryotic diversity (Fig. 4A, C and E).

The microbial communities in the lake maintained relatively high species evenness throughout the study (Fig. 4G and H), despite fluctuations in taxonomic richness (Fig. 4A–F) and seasonal changes in the chemical/physical environment, with one exception: a pronounced decrease in eukaryotic evenness during the *P. parvum* bloom in Lebanon Pool. The Simpson index of diversity reflects species richness, but is strongly affected by the relative abundances of taxa within an assemblage. With the exception of the months of the biological disturbance event, our data suggest that the microbial communities in this system are structured to maintain high evenness despite over twofold changes in total richness. Examination of community similarities among the monthly samples using the Bray–Curtis and Jaccard indices demonstrated significant annual cycles for the eukaryotic and bacterial assemblages at both sampling sites. The yearly trajectory of the months in the MDS plots indicated a statistically supported cyclical seasonal progression, and a return of community structure to the approximate starting point at each location at the end of the year-long study (Fig. 5A–D). Similar patterns were detected with both the Bray–Curtis (which takes into account shared taxa and relative abundances of OTUs), and the Jaccard (which considers presence/absence and thus common and rare taxa are given equal weight) indicating that both patterns of community membership and patterns in the catalogue of taxa present were affected in parallel by seasonal environmental forcing factors. The conserved pattern of cyclical change implies an active seasonal response among the less dominant taxa in the microbial communities in parallel with the more numerous members.

Microbial community response to an ecosystem disruptive harmful algal bloom

The *P. parvum* bloom that occurred in Lebanon Pool and not Wilson Creek created the conditions for a natural experiment in which to explore the effects of an ecosystem disruptive harmful algal bloom on eukaryotic and bacterial microbial community structure. The overall effect of this harmful alga on food web structure derives from the direct effects of toxins produced by this species on a wide range of organisms (Skovgaard and Hansen, 2003), as well as its well-characterized phagotrophic ability (Tillmann, 1998). Toxins produced by *P. parvum* have been shown to negatively affect phototrophic and heterotrophic species of the microbial plankton, and its phagotrophic capabilities allow the ingestion of taxa ranging in size from bacteria to zooplankton that can be significantly larger than the alga (Tillmann, 2003).

The *P. parvum* bloom in Lebanon Pool had a surprisingly limited impact on microbial taxonomic richness. The number of observed eukaryotic OTUs per month was lower over the time span of the bloom in Lebanon Pool compared with Wilson Creek (Fig. 4A), but similar estimates of total eukaryotic species richness were observed at both sites throughout the bloom using the Catchall and Chao richness estimators (Fig. 4C and E). This result presumably indicates that features of the eukaryotic assemblage structure on which these species richness estimators are based were similar at both study sites during the *P. parvum* bloom in Lebanon Pool. Negative effects were not apparent on the observed number of bacterial OTUs or total bacterial species richness estimates as a consequence of the *P. parvum* bloom (Fig. 4b, D and F). Despite the pronounced dominance of a single taxon during the bloom, a diverse suite of eukaryotic and bacterial microbes remained present as subdominant or rare taxa.

The *P. parvum* disturbance event had a strong negative effect on the taxonomic structure (depressed evenness and altered distributional trends) of the eukaryotic assemblage.

The Simpson Index of Diversity was markedly decreased in Lebanon Pool from January to April (Fig. 4G) (the timing and location of the *P. parvum* bloom; Figs 2H, 3A and B). The low evenness was not solely due to the over-dominance of *P. parvum* because low values also were obtained when *P. parvum* sequences were removed from the dataset and the dataset was re-analysed (dotted line in Fig. 4G). Low Simpson values derived using the latter dataset were relegated to the end of the *P. parvum* bloom, implying that community structure was strongly skewed towards fewer, highly dominant species and provided evidence that the effects of the bloom were radiating into the eukaryotic community selecting for physiologies that were tolerant to, or benefited from, the presence of the toxic haptophyte.

Analysis of community similarities also indicated a strong impact of the *P. parvum* bloom on both the eukaryotic and bacterial assemblages during the study (Fig. 5). The patterns in the annual trajectories of community similarities were highly consistent between the bacterial and the eukaryotic assemblages, including the divergence between Lebanon Pool and Wilson Creek during the *P. parvum* bloom. This statistically supported separation (see *Results*) was observed in both the Bray–Curtis and Jaccard plots, and suggests that the

effects of the bloom on the microbial communities were far-reaching, affecting the most abundant as well as rare eukaryotic and bacterial taxa (Fig. 5A–D). Therefore, while bacterial species richness and evenness were minimally affected by the *P. parvum* bloom, the composition of the bacterial assemblage changed in a manner that paralleled changes in the eukaryotic assemblage suggesting coordinated co-occurrences.

Episodic physical disturbance

The major rain event in the beginning of May (Fig. S1) and resulting decreases in salinity, pH and dissolved oxygen (Fig. 2B–D) led to dramatic changes in the eukaryotic and bacterial assemblages at both study sites. Community compositional changes were apparent as dramatic shifts in the May samples on the MDS plots of community similarity. The samples from May plotted closely together for the two sampling sites, but apart from all other samples, for both the bacterial and eukaryotic assemblages (i.e. low similarity to all other samples; Fig. 5A–D).

The rain event also brought an end to differences in microbial community similarities between the two sampling locations caused by the *P. parvum* bloom in Lebanon Pool, and thus acted to 'reset' the eukaryotic and bacterial communities at these locales back to a high level of similarity (albeit with community composition that was different from any other month). It would be anticipated that the rain and runoff and consequent changes in chemical/physical parameters would introduce and/or select for a unique suites of microbial taxa. Indeed, the largest increase in observed eukaryotic and bacterial OTUs between months at both locations occurred between April (pre-rain event) and May (post-rain event; Fig. 4A and B).

Conclusions

Changes in the eukaryotic and bacterial diversity (species richness and evenness), community membership and structure observed over an annual cycle at two locations within Lake Texoma reflected distinct and clearly identifiable responses to (i) seasonal environmental forcing factors, (ii) a shared episodic physical disturbance (a major rain event) and (iii) a prolonged and localized ecosystem disruptive harmful algal bloom (of the toxic alga, *P. parvum*). The latter two responses were overlaid on what appeared to be an annual cyclicity of environmental forcing that acted to 'reset' the microbial community structure close to the same starting point at the beginning and end of our year-long study. The microbial communities at the two locations were similar in any given month, suggesting a shared response to the abiotic environment. The bacterial, and eukaryotic communities were each different at the two locations during the harmful algal bloom disturbance event, which highlights the role of local processes including interactions between the eukaryotic and bacterial communities. Overall, our results were consistent with the hypothesis that environmental perturbations (physical, chemical and biological forcing factors) result in significant and rapid reassembly of the microbial community (Caron and Countway, 2009) to maintain biological community structure.

Experimental procedures

Sample collection and laboratory processing

Near-shore water samples were collected monthly from November 2008 to October 2009 at two locations (Lebanon Pool and Wilson Creek; Fig. 1) for temperature, specific conductance, chlorophyll *a*, nutrients, dissolved oxygen, pH, *Prymnesium parvum* cell counts, and DNA for subsequent sequencing of 18S and 16S rRNA genes. Wilson Creek was sampled twice on the same day in December and this additional sample served as a true replicate to examine the variance associated with sample collection and processing. The frequency of sampling for temperature, salinity, chlorophyll *a*, and *P. parvum* counts was greater than once a month for part of the year.

Temperature, salinity, pH, dissolved oxygen, nutrients, chlorophyll *a*, and *P. parvum* cell counts were collected and analysed as described in (Hambright *et al.*, 2010). Water samples of 100–300 ml were filtered onto 47 mm GF/F filters (Whatman) and stored at −80°C for later DNA extraction, PCR amplification and sequencing.

DNA was extracted from one quarter of each filter as described in (Countway *et al.*, 2007). DNA was then PCR-amplified with a cocktail of primers targeting the v9 region (~ 120 bases) of the 18S rRNA gene for eukaryotes and the v6 region (~ 90 bases) of the 16S rRNA gene for bacteria (Sogin *et al.*, 2006; Amaral-Zettler *et al.*, 2009). The 18S and 16S rRNA gene fragments were amplified for each of the 25 samples in 7–10 individual 25 μl PCR reactions each. All reactions were performed according to the following protocol: 0.45 μM cocktail of each forward and reverse primers, 1× GoTaq Flexi Reaction Buffer (Promega, Madison WI), 2.5 mM magnesium chloride, 200 mM dNTPs, 0.8 μg ml^{-1} bovine serum albumin (BSA), 2.5 Units of GoTaq Flexi DNA Polymerase (Promega) and 10 ng of total DNA. The following touchdown PCR protocol was used: 1 cycle (95°C for 2 min), 10 cycles (95°C for 30 s, 65°C-decreasing 1°C per cycle for 30 s, 72°C for 30 s), 18 cycles (95°C for 30 s, 55°C for 30 s, 72°C for 30 s), and a final extension of 72°C for 7 min. The multiple PCR reactions were pooled, cleaned and concentrated (Zymo Research). The amplicons (~ 500 ng total) were then amended with 454 adaptors and DNA barcodes according to Roche manufacturing protocols and sequenced with Roche Life Sciences 454 Titanium chemistry and platform. The 50 libraries (two genes at two locations for twelve months plus one duplicate) were multiplexed on 2 one-quarter sequencing plates. Sequence barcodes and identifier keys were removed after sequencing using Roche software (Margulies *et al.*, 2005).

Bioinformatic analyses of 18S and 16S rDNA sequences

The free software package MOTHUR v1.18.0 (http://www.mothur.org) was used for sequence processing, taxonomic identification, alignments, OTU calling and alpha diversity analyses (Schloss *et al.*, 2009). The sequences were screened for quality and only the sequences that met the following criteria were retained: an average quality score of 25 or higher, zero ambiguous bases, homopolymers with less than 8 bases, and exact matches to the proximal and distal primers (Huse *et al.*, 2007).

Plastids were removed from the bacterial dataset to avoid biasing bacterial community structure with eukaryotic plastid sequences. Each 16S sequence was compared against the 16S SILVA taxonomy reference database within MOTHUR. Sequences identified as cyanobacteria were manually checked (blastn) against the web-based NCBI (nt) database (Zhang *et al.*, 2000) to ensure they were not plastids. Removing the plastids from the 16S dataset reduced the total number of 16S sequences by approximately 10%.

The total number of high quality eukaryotic (18S rDNA) sequences/sample ranged from 7678 to 24 934 for the 25 samples analysed. Bacterial sequences/sample ranged from 3597 to 18 017. Each library in the 18S (including metazoa; approximately 5% of the sequences) and 16S (without plastids) was randomly subsampled to the month/site with the fewest number of sequences (7678, and 3597 for 18S and 16S respectively). All 25 subsampled libraries of each type were then pooled for calling OTUs.

The subsampled pooled datasets were each aligned to a reference SILVA alignment (18S rDNA or 16S rDNA). Sequences that did not align well to the V9 region of the 18S or V6 region of the 16S were removed from each dataset (average $< 2\%$ for 18S and $< 8\%$ for 16S from any library). We performed a pre-clustering step on the sequences to condense those with single base pair differences that may have been the result of sequencing errors (Huse *et al.*, 2010). OTUs were called at 97% similarity for each of the two subsampled datasets using an average neighbour method. A representative sequence (one with the smallest total distance to all the other sequences within an OTU) was extracted from each OTU and compared with the SILVA small subunit ribosomal (v108) using stand alone BLAST+. Each OTU was assigned an identification based on a best match of the representative sequence in SILVA. A best match was defined as greater than 95% sequence similarity over 95% of the query sequence. Using this approach the overall number of OTUs may be overestimated compared the number using PyroNoise as described by Schloss and colleague (2011) (see Fig. S3). However, we are confident in the overall data trends.

Statistical analyses of diversity, community structure and seasonal patterns

The Chao1 (Chao and Lee, 1992; Chao, 2005) and Catchall (Bunge, 2011) richness estimators, as well as the Simpson index of diversity (1-D) which takes into account taxonomic evenness and richness (Simpson, 1949) were computed with upper and lower 95% confidence intervals using MOTHUR for the subsampled 18S libraries and 16S libraries without plastids. The Simpson index of diversity also was calculated for the 18S Lebanon Pool dataset following removal of all sequences attributed to *Prymnesium*, and the dataset again subsampled and the indices recalculated. All linear regressions, fitting of trend lines and simple correlation analyses were computed and plotted in SigmaPlot v11.

The multivariate software package PRIMER v6.1.7 was used to compute Bray–Curtis (Bray and Curtis, 1957) and Jaccard (Jaccard, 1908) similarity values between each month and location, perform permutation-based statistical tests, and build cluster and non-metric MDS plots (Clarke, 1993; Clarke and Warwick, 2001). Bray–Curtis similarity matrices were built for each dataset (18S and 16S libraries) pretreated in the following two manners: (i) Subsampled (as noted above), with OTU abundances converted to relative abundances; (ii) Subsampled, with OTU abundances converted to relative abundances and square root transformed. In order to assess the similarity of these data treatments, Spearman correlations between the matrices were computed using second stage MDS. Comparison of the square-root transformed and untransformed libraries also yielded high Spearman correlations (0.97) for both datasets. Based on these high correlations, the remainder of the analyses were conducted using the subsampled, untransformed OTUs converted to relative abundances. Jaccard similarity matrices were built for each subsampled dataset (18S and 16S libraries).

The ANOSIM procedure was used to test for differences in community similarities between the two locations, and between the months within each location. We treated the *P. parvum* bloom (documented by microscopical cell counts in Lebanon Pool from January to April) as a factor and used ANOSIM to test for differences in community similarities between the locations during bloom and non-bloom conditions.

Cluster diagrams were constructed from the Bray–Curtis and Jaccard similarity matrices, and the hierarchical clusters of samples were organized based on group average scores. The SIMPROF permutation test was used to identify statistically significant samples within the clusters. The same similarity matrices used to build the dendrograms were converted into ranks and then used to construct two-dimensional MDS plots with a Kruskal fit scheme of 1 and 25 restarts. The placement of the samples is considered a good fit when the stress is less than 0.1. The annual trajectory of the samples was depicted using arrows to connect sequential months for each location on the plots.

Repeating patterns at each location within the eukaryotic or bacterial communities were examined using Spearman correlations to a cyclical model matrix. Relationships between the 18S and 16S datasets and. the Bray–Curtis and Jaccard indices were tested by computing Spearman correlations between the similarity matrices. These correlations were statistically tested using the RELATE permutation test (Clarke and Warwick, 2001; Clarke *et al.*, 2006).

The global BEST procedure was used to identify the environmental variables or sets of variables that best matched the patterns detected in the biological community similarity analyses, it returns a test statistic and a permuted *P*-value (Clarke *et al.*, 2008). The BEST routine was performed with the

following environmental parameters that were square root transformed and normalized: temperature, PSU, dissolved oxygen, pH, total dissolved nitrogen, total nitrogen, total dissolved phosphorous, total phosphorous, and both total and dissolved molar N : P ratios. The BEST routine was performed a second time without the biological variables of cell counts, chlorophyll and dissolved oxygen.

Acknowledgements

Funding was provided by a grant from the Oklahoma Department of Wildlife Conservation through the Sport Fish Restoration Program (grant F-61-R) to K.D.H. Oklahoma Mesonet data are provided courtesy of the Oklahoma Mesonet, which is jointly operated by Oklahoma State University and the University of Oklahoma. Continued funding for maintenance of the network is provided by the taxpayers of Oklahoma. The authors would like to thank James D. Easton, Anne C. Easton, and Richard Zamor for field measurements, collecting samples, microscope counts.

All sequences are located in the NCBI short read archive under the BioProject PRJNA195945.

References

Allison, S.D., and Martiny, J.B.H. (2008) Resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci USA* **105:** 11512–11519.

Amaral-Zettler, L.A., Gomez, F., Zettler, E., Keenan, B.G., Amils, R., and Sogin, M.L. (2002) Microbiology: eukaryotic diversity in Spain's River of Fire. *Nature* **417:** 137–137.

Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., and Huse, S.M. (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE* **4:** e6372.

Behnke, A., Engel, M., Christen, R., Nebel, M., Klein, R.R., and Stoeck, T. (2011) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environ Microbiol* **13:** 340–349.

Bouvier, T.C., and del Giorgio, P.A. (2002) Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries. *Limnol Oceanogr* **47:** 453–470.

Bray, J., and Curtis, J. (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* **27:** 325–349.

Bunge, J. (2011) Estimating the number of species with CatchAll. Paper presented at: Proceedings of Pacific Symposium on Biocomputing (Kohala Coast, Hawaii).

Buskey, E.J., Liu, H., Collumb, C., Guilherme, J., and Bersano, J.G.F. (2001) The decline and recovery of a persistent Texas brown tide algal bloom in the Laguna Madre (Texas, USA). *Estuaries* **24:** 337–346.

Caron, D.A., and Countway, P.D. (2009) Hypotheses on the role of the protistan rare biosphere in a changing world. *Aquat Microb Ecol* **57:** 227–238.

Chao, A. (2005) *Species Estimation and Applications*, Vol. **12**. New York, USA: Wiley.

Chao, A., and Lee, S.M. (1992) Estimating the number of classes via sample coverage. *J Am Stat Assoc* **87:** 210–217.

Clarke, K.R. (1993) Non-parametric multivariate analyses of changes in community structure. *Aust J Biol Sci* **18:** 117–143.

Clarke, K.R., and Warwick, R.M. (2001) *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation*, 2nd edn. Plymouth, UK: PRIMER-E.

Clarke, K.R., Somerfield, P.J., and Chapman, M.G. (2006) On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages. *J Exp Mar Biol Ecol* **330:** 55–80.

Clarke, K.R., Somerfield, P.J., and Gorley, R.N. (2008) Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *J Exp Mar Biol Ecol* **366:** 56–69.

Connell, J.H. (1978) Diversity in tropical rain forests and coral reefs. *Science* **199:** 1302–1310.

Countway, P.D., Gast, R.J., Dennett, M.R., Savai, P., Rose, J.M., and Caron, D.A. (2007) Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the western N. Atlantic (Sargasso Sea and Gulf Stream). *Environ Microbiol* **9:** 1219–1232.

Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C., *et al*. (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J* **5:** 1344–1356.

Edgcomb, V.P., Kysela, D.T., Teske, A., Gomez, A., and Sogin, M.L. (2002) Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proc Natl Acad Sci USA* **99:** 7658–7662.

Estes, J.A., Terborgh, J., Brashares, J.S., Power, M.E., Berger, J., Bond, W.J., *et al*. (2011) Trophic downgrading of planet Earth. *Science* **333:** 301–306.

Evardsen, B., and Imai, I. (2006) The ecology of harmful flagellates within Prymnesiophyceae and Raphidophyceae. In *Ecology of Harmful Algae*. Granéli, E., and Turner, J. (eds). Berlin, Germany: Springer-Verlag, pp. 67–79.

Fistarol, G., Legrand, C., and Granéli, E. (2003) Allelopathic effect of *Prymnesium parvum* on a natural plankton community. *Mar Ecol Prog Ser* **255:** 115–125.

Floder, S., and Sommer, U. (1999) Diversity in planktonic communities: an experimental test of the intermediate disturbance hypothesis. *Limnol Oceanogr* **44:** 1114–1119.

Fuhrman, J.A. (2009) Microbial community structure and its functional implications. *Nature* **459:** 193–199.

Gihring, T.M., Green, S.J., and Schadt, C.W. (2012) Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environ Microbiol* **14:** 285–290.

Gilbert, J.A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., *et al*. (2009) The seasonal structure of microbial communities in the Western English Channel. *Environ Microbiol* **11:** 3132–3139.

Gobler, C.J., and Sunda, W.G. (2012) Ecosystem disruptive algal blooms of the brown tide species, *Aureococcus*

anophagefferens and *Aureoumbra lagunensis*. *Harmful Algae* **14:** 36–45.
Gobler, C.J., Lonsdale, D.J., and Boyer, G.L. (2005) A review of the causes, effects, and potential management of harmful brown tide blooms caused by *Aureococcus anophagefferens* (Hargraves et Sieburth). *Estuaries* **28:** 726–749.
Hambright, K.D., Zamor, R.M., Easton, J.D., Glenn, K.L., Remmel, E.J., and Easton, A.C. (2010) Temporal and spatial variability of an invasive toxigenic protist in a North American subtropical reservoir. *Harmful Algae* **9:** 568–577.
Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* **8:** R143.
Huse, S.M., Welch, D.M., Morrison, H.G., and Sogin, M.L. (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* **12:** 1889–1898.
Jaccard, P. (1908) Nouvelles recherches sur la distribution florale. *Bull Soc Amis Sci (Med) (Poznan)* **44:** 223–270.
Kent, A.D., Yannarell, A.C., Rusak, J.A., Triplett, E.W., and McMahon, K.D. (2007) Synchrony in aquatic microbial community dynamics. *ISME J* **1:** 38–47.
Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12:** 118–123.
Langenheder, S., Kisand, V., Wikner, J., and Tranvik, L.J. (2006) Salinity as a structuring factor for the composition and performance of bacterioplankton degrading riverine DOC. *FEMS Microbiol Lett* **45:** 189–202.
Lindström, E.S., and Langenheder, S. (2011) Local and regional factors influencing bacterial community assembly. *Environ Microbiol* **4:** 1–9.
Lopez-Garcia, P., Philippe, H., Gail, F., and Moreira, D. (2003) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc Natl Acad Sci USA* **100:** 697–702.
McPherson, R.A., Fiebrich, C.A., Crawford, K.C., Elliott, R.L., Kilby, J.R., Grimsley, D.L., *et al*. (2007) Statewide monitoring of the mesoscale environment: a technical update on the Oklahoma Mesonet. *J Atmos Oceanic Technol* **24:** 301–321.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., *et al*. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376–380.
Martiny, J.B.H., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., *et al*. (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4:** 102–112.
Medinger, R., Nolte, V., Pandey, R.V., Jost, S., Ottenwalder, B., Schlotterer, C., and Boenigk, J. (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Mol Ecol* **19** (Suppl. 1)**:** 32–40.
Michaloudi, E., Moustaka-Gouni, M., Gkelis, S., and Pantelidakis, K. (2008) Plankton community structure during an ecosystem disruptive algal bloom of *Prymnesium parvum*. *J Plankton Res* **31:** 301–309.
Pedros-Alio, C. (2006) Marine microbial diversity: can it be determined? *Trends Microbiol* **14:** 257–263.
Reynolds, C.S., Padisak, J.A., and Sommer, U. (1993) Intermediate disturbance in the ecology of phytoplankton and the maintenance of species diversity: a synthesis. *Hydrobiologia* **249:** 183–188.
Roelke, D.L., Grover, J.P., Brooks, B.W., Glass, J., Buzan, D., Southard, G.M., *et al*. (2010) A decade of fish-killing *Prymnesium parvum* blooms in Texas: roles of inflow and salinity. *J Plankton Res* **33:** 243–253.
Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.
Schloss, P.D., Gevers, D., and Westcott, S.L. (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* **6:** e27310.
Scholin, C., Gulland, F., Doucette, G.J., Benson, S., Busman, M., Chavez, F.P., *et al*. (2000) Mortality of sea lions along the central California coast linked to a toxic diatom bloom. *Nature* **403:** 80–84.
Shade, A., Read, J.S., Youngblut, N.D., Fierer, N., Knight, R., Kratz, T.K., *et al*. (2012) Lake microbial communities are resilient after a whole-ecosystem disturbance. *ISME J* **6:** 2153–2167.
Simpson, E.H. (1949) Measurement of diversity. *Nature* **163:** 688.
Skovgaard, A., and Hansen, P.J. (2003) Food uptake in the harmful alga *Prymnesium parvum* mediated by excreted toxins. *Limnol Oceanogr* **48:** 1161–1166.
Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., *et al*. (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103:** 12115–12120.
Sommer, U., Gliwicz, Z.M., Lampert, W., and Duncan, A. (1986) The PEG-model of seasonal succession of planktonic event in fresh waters. *Arch Hydrobiol* **106:** 433–471.
Sunda, W.G., Graneli, E., and Gobler, C.J. (2006) Positive feedback and the development and persistence of ecosystem disruptive algal blooms. *J Phycol* **42:** 963–974.
Tillmann, U. (1998) Phagotrophy by a plastidic haptophyte, *Prymnesium patelliferum*. *Aquat Microb Ecol* **14:** 155–160.
Tillmann, U. (2003) Kill and eat your predator: a winning strategy of the planktonic flagellate *Prymnesium parvum*. *Aquat Microb Ecol* **32:** 73–84.
Vigil, P., Countway, P.D., Rose, J., Lonsdale, D.J., Gobler, C.J., and Caron, D.A. (2009) Rapid shifts in dominant taxa among microbial eukaryotes in estuarine ecosystems. *Aquat Microb Ecol* **54:** 83–100.
Wetzel, R.G. (2001) *Limnology: Lake and River Ecosystems*, Vol. **3rd**. Philadelphia, PA, USA: Academic Press.
Zamor, R.M., Glenn, K.L., and Hambright, K.D. (2012) Incorporating molecular tools into routine HAB monitoring programs: using qPCR to track invasive *Prymnesium*. *Harmful Algae* **15:** 1–7.
Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7:** 203–214.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. An annual cycle of daily rainfall at the Madill OK, station (Black Circles). Rainfall data are taken from: http://www.mesonet.org.

Fig. S2. Cluster Diagrams of eukaryotic (A, B) and bacterial (C, D) community structure in Lebanon Pool (black triangles), Wilson Creek (grey circles) and a duplicate sample from Wilson Creek in December (grey x) based on group average clustering of Bray–Curtis similarity values (A and C) and Jaccard similarity values (B and D) calculated from non-transformed absolute OTU abundances. Dotted lines represent clusters of samples that were not significantly different from one another on the basis of the SIMPROF permutation test. Vertical grey dashed lines represent the level of similarity for the December Wilson Creek replicates.

Fig. S3. A comparison of the number of observed OTUs called at various dissimilarity cut-offs (0% to 99%) for the 16s January Lebanon Pool sample using the procedure outlined in this paper (black) compared to the PyroNoise procedure (grey). The sample was not subsampled and contained 7408 sequences.